

Program Evaluation through Kirkpatrick's Framework

Ömer Gökhan ULUM

Research Assistant
English Language Teaching Department, Hakkari University
Hakkari, Turkey

Abstract

Being purposeful, systematic and carefully implemented, evaluation is a continuous process and it is performed as the very basic part of the program activities to attain data to conclude if there is a need to make changes or eliminations, or accept something in it. Program evaluation is a kind of examination in social research field and it checks the sufficiency of educational programs. The broadest purpose of evaluation is to contribute judgments about the worth of an evaluated program or to point to the value of the program or just a section of it. The evaluators choose an evaluation model among several ones, each of which has its own characteristics or way of approaching the evaluation. One of these models is Four-level Evaluation Model by Kirkpatrick. This study was conducted based on the Document Analysis Technique by means of inquiring Kirkpatrick's framework from various sources of academic books and articles. From the analysis, one can conclude that Kirkpatrick's four-level model of program evaluation is one of the mostly employed models by the program evaluators. Besides, this study offers a documented data of how Kirkpatrick's framework that is easy to be implemented functions and what its features are.

Keywords:

Program, program evaluation, Kirkpatrick's four level evaluation model.

Introduction

Smith (1989) (as cited in Owen, 1999) defines a program as: 'a set of planned activities directed toward bringing about specified change(s) in an identified and identifiable audience' (p.47). According to Demirel (2007), education program consists of such elements as the list of topics, the contents of the course, the programming of the tasks, the list of educational materials, the arrangement of the courses, the group of objective behaviors, everything taught inside and outside school and everything planned by school staff. He also states that the employment of a program takes various stages and the successful execution of a program is realized by means of presenting the outline of program stages and development. The implementation and evaluation of the program form the final stages of this outline as Demirel (2007) mentions as well. Furthermore, the concept of program in education is categorized under some titles as education program, training program, course program, unit and lesson plan, whereas the education program is the broadest term among them (Yüksel and Sağlam, 2014, p.6). Uşun (2012) states that various program definitions have been made in the related literature. However, he defines program as a followed route which provides the related aim, the content, the order of the content, and how, where, when

and with whom this content will be executed.

Evaluation is a process that we perform to attain data to conclude if there is a need to make changes or eliminations or to accept something in the curriculum (Ornstein and Hunkins, 1998). Wall (2014) describes evaluation as a purposeful, systematic, and careful collection and analysis of information that we use with the aim of documenting the effectiveness and impact of programs, setting up liability and identifying areas in need of change and improvement. He also puts forward that evaluation is a continuous event which isn't conducted only once, and which ought to be an integral and integrated section of the program activities. Properly formed, considerably and accurately carried out evaluations can supply significant information to report the outcomes of the program and lead us toward parts where changes might be required (p.19).

Harris (1968) identifies evaluation as a systematic process to determine the worth, strength, sufficiency or allure of something with respect to specific criteria and goals. Program evaluation is the process of judging the worth of a program and this judgment is shaped by comparing evidence as to what the program is with criteria about what the program should be (Steele, 1970). It is clear that evaluations are capable of specifying the unintended effects of

programs, which can affect overall assessments of programs accordingly (Mc David, Huse and Hawthorn, 2013, p. 3). Uşun (2012) describes program evaluation as a decision process as to accuracy, authenticity, sufficiency, convenience, productivity, effectiveness, utility, success and executability of a developed program by means of employing scientific research processes based on systematic data collection and analysis. The broadest purpose of evaluation is to contribute judgments about the worth of whatever is being evaluated or to conclude the value of the program or some part of it (Fitzpatrick, Sanders and Worthen, 2004).

Mc David, Huse and Hawthorn (2013) state that program evaluators are expected to come up with ways of announcing whether the program attained its aims—whether the planned outcomes were grasped. They also refer that there aren't any program evaluations which can be achieved without some important elements such as the evaluator's own experiences, expectations, values and beliefs. Luo (2010, p.47) refers to the role of the evaluators as discussion among evaluation theorists about the definite roles of an evaluator reflects their distinct attitudes on other main perspectives such as;

- the value of evaluation (descriptive vs. prescriptive),
- the methods of evaluation (quantitative vs. qualitative),
- the use of evaluation (instrumental vs. enlightenment),
- the purpose of evaluation (summative vs. formative).

Stake (1999) states that referring to the quality of the evaluand is among the responsibilities of competent evaluators. There are six categories of evaluand, the object of evaluation, or that which is being evaluated, as programs, policies, performances, products, personnel and proposals (Leavy, 2014). The evaluand may be misrepresented as a result of a single perspective being featured (Stake, 1999).

Evaluation standards cover criteria to guide evaluators, to evaluate a conducted program evaluation or to present supported information to the authorities in terms of reliability and validity of the evaluation (Sağlam and Yüksel, 2007). Fitzpatrick, Sanders and Worthen (2004, p.445) state the evaluation standards as utility standards which are aimed to assure that an evaluation will aid the information needs of its expected users; feasibility standards which are aimed to assure that an evaluation will be realistic, reasonable, strategic, practical and economical; propriety standards which are aimed to assure that an evaluation will be achieved officially, ethically, and with regard of the

prosperity of those included in the evaluation as well as those influenced by its results; accuracy standards which are aimed to assure that an evaluation will disclose and transmit technically satisfactory information about the components or features that decide the worth or merit of the program being evaluated.

Bass (2001) states that extensive program evaluation improvements in terms of approaches took place in the last half of the 20th century and our age is a beneficial time for evaluators to analytically assess their program evaluation approaches and also to determine which ones are most satisfying for constant utilization and additional improvement. Efficient program evaluation is more than gathering, analyzing, and supplying data as it ensures collecting and using information to learn about programs continuously and also to develop them (W.K. Kellogg Foundation Logic Model Development Guide, 2004). Program evaluation models form the basis of the needed logic to analyze the outcomes of the program (Uşun, 2012). The evaluators follow different approaches and models in collecting and analysing data when evaluating the program. Furthermore, the evaluators' level of knowledge and skills of evaluation, adopted evaluation theories and philosophical values construct their program evaluation approaches (Yüksel and Sağlam, 2014). In this paper, Kirkpatrick's Evaluation Model or its four level evaluation framework is described in detail.

The Aim of the Study

The aim of this study is to present detailed perspectives as to one of the mostly used evaluation models, Kirkpatrick's four level evaluation model, by means of document analysis technique. With this in mind, this study tries to enlighten the evaluators' mind referring to the framework of widely used and easily implementable Kirkpatrick's evaluation model.

Research Method

This study is a qualitative research having resource to the document analysis technique. In other words, document analysis was used as the method of data collection and analysis in this study. In the document analysis technique, the already being records, documents or other kinds of resources are investigated and the data are acquired (Karasar, 2012). Peute (2013) states that document analysis is a form of qualitative research in which documents are illustrated by the researcher to give voice and meaning around an assessment topic.

Kirkpatrick's Evaluation Model

Kirkpatrick's four level evaluation model is extensively employed to evaluate the effectiveness of educational programs (Gill and Sharma, 2013). Donald Kirkpatrick formulated the four levels of evaluation and each level

presents an order of steps to evaluate educational programs (Meghe, Bhise and Muley, 2013). *Reaction* level evaluates the approach of the student towards the program; *learning* level evaluates the knowledge achieved by the sample population having been exposed to the education; *behavior* level measures how properly the knowledge achieved is put into use by trainees; *results* level measures how appropriately the major aim of the education is attained (Alturki and Aldraiweesh, 2014). Namely, Gill and Sharma (2013) define the levels as reaction evaluates how the students feel about the program, learning evaluates the amount of learning achieved, behavior is the degree of behavior change and finally results are the real gains of the educational program. According to the model each level is significant and is in contact with the next level (Gill and Sharma, 2013). The Kirkpatrick four-level evaluation model has acted as the fundamental regulating scheme for educational evaluations for about more than 40 years and there is no questioning about the model's having made significant supplement for educational evaluation practices (Bates and Coyne, 2005). However, in their study, Bates and Coyne (2005) also mention that the insufficiency of Kirkpatrick's 4-level model to contain application of crucial circumstantial input variables in educational evaluation conceals the actual complexities of the educational progress. That's to say, they put forward that the trouble with employing the four level model of Kirkpatrick is that though it might supply some gainful data as to program results, when evaluation is confined to educational consequences no data about why education was or was not efficient is brought about. Frye and Hemmer (2012) refer to the model's main educational evaluation aids as the comprehensibility of its concentration on program results and its crystal-clear explanation about the results beyond basic student gladness. Kirkpatrick advised collecting information to specify four hierarchical levels of program results: (1) student contentment or responsiveness for the program; (2) measurements of acquisition such as achieved knowledge, developed skills and behaviours as a result of the program; (3) differences in student's behaviour in the atmosphere in which they are educated; and as a consequence (4) the program's last outcomes in its broader context (Frye and Hemmer, 2012). Furthermore, in the study which Frye and Hemmer (2012) conducted in 2012, they indicate that to understand student reactions to the program, evaluators should choose the wished reactions such as learners' contentment and ask the students' opinions about the education program. For instance, the students may be asked if they sensed the program was beneficial for their learning or not, according to what Frye and Hemmer (2012) mention. They also state that the following Kirkpatrick level necessitates the evaluator to specify what participants have acquired in the process of the program. The level three concentrates on student behavior in the context for which

they were educated; for instance post-graduate students' adoption of the program's knowledge and skills may be seen in their setting of the practice and equated with the asked standard to gather clue of the level three (Frye and Hemmer, 2012). They sum up the Kirkpatrick's four level as an evaluation level concentrating on student outcomes noticed after a proper duration in the program's broader context: the program's influence on such aspects as outcomes, savings, performance, etc. Kirkpatrick's framework is described in detail in the following sections.

Reaction

Reaction is Kirkpatrick's first level of evaluation, which evaluates how the participants living the learning experience perceive the action (Kirkpatrick, 1998). Nelson and Dailey (1999) put forward that reaction is mainly acquired at the final stage of education by basically asking the participants, for instance; "How did the education feel to you?". Generally formed as a survey or questionnaire, participants hint this level as "happy sheets" or "feel-good measure" and an organized way as to participants' respond to the program could contain basic questions such as (Nelson and Dailey, 1999):

- Is your work group excited about the recognition program?
- Did the program describe how and why you should recognize others?
- Are the program guidelines clear and communicated well?
- Is the nomination and award process simple to use?
- Do you like the merchandise or activities provided as re-wards for the program?
- How is it better than the previous program or activity?
- What is your favorite part of the program?
- Are there areas for improvement?

Kirkpatrick (1998) states the aim of measuring reaction is to guarantee that participants are motivated and involved in learning. He shows the implementation guidelines of reaction level as in the following:

- Determine what you want to find out.
- Design a form that will quantify reactions.
- Encourage written comments and suggestions.
- Attain an immediate response rate of 100%.
- Seek honest reactions.
- Develop acceptable standards.

- Measure reactions against the standards and take appropriate action.
- Communicate the reactions as appropriate.

Learning

Kirkpatrick's second level of evaluation is *learning*. Kirkpatrick describes this level as the scope in which participants in the program alter approaches, enhance knowledge, or develop skills in lieu of the program (Kirkpatrick, 1998). Kirkpatrick's Level 2 evaluation measures the acquired knowledge a student has achieved by joining the training (DOL Connecting Network and Career Development, 2011). Learning evaluates the amount of participants' achieved experiences, attitudes, and principles involved in the education process (Lynch, Akridge, Schaffer and Gray, 2006). We can evaluate if specific abilities or awareness levels have been transformed into more developed ones as a result of the program and some other measurable acquisitions contain the followings as well (Nelson and Dailey, 1999):

- Using formal, informal and day-to-day recognition
- Knowing how to praise publicly
- Timing the recognition appropriately
- Writing a persuasive nomination for an employee award
- Knowing what forms of recognition work well for different types of performance

As mentioned, Kirkpatrick describes learning as the point at which those taking part in the program reach by means of shifted attitudes, raised knowledge and promoted skills as a result of joining the program (Nelson and Dailey, 1999). Application of this new knowledge, skills, or attitudes is not evaluated at this level, though (Kirkpatrick, 1998). What Kirkpatrick (1998) also refers about the implementation guidelines of Learning Level follows as:

Use a control group, if feasible.

- Evaluate knowledge, skills, or attitudes both before and after training.
- Use a paper and pencil test to measure knowledge and skills.
- Use a performance test to measure attitudes.
- Attain a response rate of 100%.
- Use the results of the evaluation to take appropriate action

Behavior

Kirkpatrick's third level of evaluation is *behavior*. This level refers to "To what degree do the learners apply what they have learnt during education?" (Kirkpatrick, 2011). That's to say, behavior level points out whether the participants are really employing what they have acquired during the program (Schumann, Anderson, Scott and Lawton, 2001). Although learning has taken place, it doesn't mean that this learning transforms into new behavior in real life (Nelson and Dailey, 1999). Behavior evaluation suggests that learners apply the pre-learnt items afterwards and change their behaviors as a result, and this might be instantly or much time after the education process, based on the position (Topno, 2012). Third level makes us conclude whether alterations in behavior have happened as a result of the program, and also Kirkpatrick points out the necessity of having data on the 1st and the 2nd levels to clarify the outcomes of the 3rd level evaluation (McLean and Moss, 2003). According to what McLean and Moss (2003) clarify if the behavior change does not appear, it is convenient to decide whether this is because of the participant's discontentment with the 1st level or lack of success in terms of the aims of the 2nd level, or whether the shortage of change in behavior is because of some other reasons like a lack of desire, aid or opportunity. Implementation guidelines of this level are as follow (Kirkpatrick, 1998):

- Use a control group, if feasible.
- Allow enough time for a change in behavior to take place.
- Survey or interview one or more of the following groups: trainees, their bosses, their subordinates, and others who often observe trainees' behavior on the job.
- Choose 100 trainees or an appropriate sampling.
- Repeat the evaluation at appropriate times.
- Consider the cost of evaluation versus the potential benefits.

Results

Results is the fourth level of evaluation in Kirkpatrick's Framework. J. Kirkpatrick (2009) and W. Kirkpatrick (2009) state that *Results Level* can be referred as to what point aimed outcomes occur as a consequence of the outcomes of the learning activity and following reinforcement. The *fourth level* or *results level* is the most challenging part to evaluate adequately and this level describes results to contain an organization's ability to learn, alter, and improve in agreement with its specified objectives (McNamara, Joyce and O'hara, 2010). "What impact has the

change produced on the organization?"(Monaco, 2014). Although we have just evaluated the initial three levels of a program, we are still unaware of what influence the program has on the institution (Nelson and Dailey, 1999). Kirkpatrick (1998) states that results mean the scale at which the institution's output has developed in lieu of the program (Schumann, Anderson, Scott and Lawton, 2001). This level means the hardest educational outcome to determine and as well as specifying the extent to which education makes a change in specific outcomes (Barbee and Antle, 2008). The objective of Kirkpatrick's 4th level evaluation is to determine organizational outcomes in terms of performance, developments and benefits as well (Kaufman, Keller and Watkins, 1995). The aim of the 4th level of evaluation is also to measure the influence of the arranged event on the institution's goals. This should obviously show the student's ability to perform more successfully as a result of the education conducted (Dhliwayo and Nyanumba, 2014). Implementation guidelines of this level are as follow (Kirkpatrick, 1998):

- Use a control group, if feasible.
- Allow enough time for results to be achieved.
- Measure both before and after training, if feasible.
- Repeat the measurement at appropriate times.
- Consider the cost of evaluation versus the potential benefits.
- Be satisfied with the evidence if absolute proof isn't possible to attain.

Conclusion

Program evaluation is the most significant aspect of education and it is a subject which has been much talked over but superficially employed (Topno, 2012). With this in mind, the aim of this article has been to analyze the Kirkpatrick's framework as an evaluation tool. Learning something from an evaluation or about it generally makes us alter our mental models or think again about our hypothesis or beliefs and improve recent comprehensions about our program evaluation processes (McNamara, Joyce and O'hara, 2010). Educational programs are simply concerning with alteration: altering students' knowledge, approach, or abilities; altering educational structures; improving educational leaders; and etc. (Frye and Hemmer, 2012). The evaluation model that we select is extensively affected by our philosophy of evaluation, though such elements as resources, time and specialization in the field also affect the employed procedures. Many program evaluation professionals are in the view that there is no solely best model, though (McNamara, Joyce and O'hara, 2010). Furthermore, in lieu of this, it is a need for the program evaluator to choose a model which responds to the

requirements of a case to form proper evaluation findings to evaluate a program's merits, worth and value as McNamara, Joyce and O'hara (2010) state as well. Arthur, Bennett, Edens and Bell (2003) employed the Kirkpatrick's framework in their study as it was theoretically the most convenient for their objectives. They refer to the Kirkpatrick's framework as inquiries about the impact of educational programs are generally pursued by questioning, "Effective in terms of what? Reactions, learning, behavior, or results?" Kirkpatrick's four-level model of program evaluation is mostly employed model and the four levels measure the followings (Austrac e-learning, 2008):

Level 1: reaction of student - what students thought and felt about the training (reaction to training)

Level 2: learning - the resulting increase in students' knowledge or capability (achievement of learning)

Level 3: behavior - extent of behavior and capability improvement and implementation/application (application of learning)

Level 4: results - effects on the business or environment resulting from the trainee's performance (organizational effectiveness).

As every evaluation level analyses the sufficiency of the program from a different aspect, each level of four is complementary and through employing all four levels, we achieve a more total frame for the sufficiency of the program (Schumann, Anderson, Scott and Lawton, 2001). Bates (2004) asks the questions "Are we doing the right thing, and are we doing it well?" to learn about the four level evaluation model of Kirkpatrick. Then, he answers the first question 'are we doing the right thing?', by stating that the simplicity and popularity of Kirkpatrick's model can be attributed to the answer. When it comes to second question, he puts forward that the limitations of Kirkpatrick's model may put barriers in front of us and employing the model may be risky for clients or stakeholders. Kirkpatrick model is the commonly employed model at reaction level, however what should be the chief indicator at this level and other levels is not described well (Topno, 2012). However, when evaluators start their search for program evaluation, they generally get closer to one of the most famous evaluation scientists, Donald Kirkpatrick (Bishop, 2010).

References

- Alturki, U. & Aldraiweesh, A. (2014). Assessing Effectiveness Of E-training Programs Based On Kirkpatrick's Model. Texas, The Clute Institute International Academic Conference.

- Arthur, W., Bennett, W., Edens, P.S. & Bell, S.T. (2003). Effectiveness of Training in Organizations: A Meta-Analysis of Design and Evaluation Features. *Journal of Applied Psychology*, 88(2), 234-245.
- Austrac e learning (2008). Using training evaluation for AML/CTF program monitoring. AML/CTF Programs. Commonwealth of Australia. Retrieved from www.austrac.gov.au
- Barbee, A.P. & Antle, B.F. (n.d.). Recommendations and Suggested Models for Colorado's Court Improvement Program Training Evaluation System. Retrieved from <https://cip-cop.icfwebsiteservices.com/>
- Bass, J. (2001). Evaluation Models. *A Publishing*, 89, 7-98.
- Bates, R. (2004). A critical analysis of evaluation practice: the Kirkpatrick model and the principle of beneficence. *Elsevier*, 27, 341-347.
- Bates, R. & Coyne, T. H. (2005). Effective Evaluation of Training: Beyond the Measurement of Outcomes. *Institute of Education Sciences (ERIC)*, 16(1), 371-378.
- Bishop, B. (2010). The Amalgamated Process for Evaluation (APE): The best of Kirkpatrick, Brinkerhoff, Dessinger & Moseley, and Phillips.
- Demirel, Ö. (2007). Eğitimde Program Geliştirme. Ankara, Pegem A Yayıncılık.
- Dhliwayo, S. & Nyanumba, L.K. (2014). An evaluation of an on the job training program at a UK based public health care company. *Problems and Perspectives in Management*, 12(2), 164-172.
- DOL Connecting Learning and Career Development. (2011). Best Practices in Instructional Design for Web-based Training. Washington, DC.
- Fitzpatrick, J., L., Sanders, J.R., Worthen, B., R. (2004). Program evaluation. Alternative approaches and practical guidelines. New York, Pearson.
- Frye, A.V. & Hemmer, P.A. (2012). Program evaluation models and related theories: AMEE Guide No. 67. Web Paper Amee Guide, 34 (e288–e299).
- Gill, M. & Sharma, G. (2013). Evaluation of Vocational Training Program from the Trainees' Perspective: An Empirical Study. *Pacific Business Review International*, 6(5), 35-43.
- Harris, W. (1968). The Nature and Function of Educational Evaluations. *Peabody Journal of Education*, 46(2), 95-99.
- Karasar, N. (2012). Bilimsel Araştırma Yöntemi. Ankara: Nobel Yayıncılık.
- Kaufman, R., Keller, J. & Watkins, R. (1995). What works and what doesn't: Evaluation beyond Kirkpatrick. *P&I*, 35(2), 8-12.
- Kirkpatrick, D.L. (1998). Another look at evaluating training programs. Alexandria, VA: American Society for Training & Development.
- Kirkpatrick, D. L. (1998). Evaluating Training Programs: The Four Levels (2nd Ed.). San Francisco, Berrett-Koehler.
- Kirkpatrick, J. & Kirkpatrick, W. (2009). The Kirkpatrick Model: Past, Present and Future. Chief Learning Officer. 20-55.
- Kirkpatrick, W. (2011). Training On Trial. [PowerPoint slides]. Kirkpatrick Partners, LLC.
- Leavy, P. (2014). The Oxford Handbook of Qualitative Research. USA, Oxford Library of Psychology.
- Luo, H. (2010). The Role for an Evaluator: A Fundamental Issue for Evaluation of Education and Social Programs. *International Education Studies*, 3(2), 42-50.
- Lynch, K., Akridge, J.T., Schaffer, S.P. & Gray, A. (2006). A Framework for Evaluating Return on Investment in Management Development Programs. *International Food and Agribusiness Management Review*, 9(2), 54-74.
- McDavid, J.C., Huse, I. & Hawthorn, L.R.L. (2013). Program Evaluation and Performance Measurement. An Introduction to Practice. Canada, SAGE Publications.
- McLean, S. & Moss, G. (2003). They're happy, but did they make a difference? Applying

- Kirkpatrick's framework to the evaluation of a national leadership program. The Canadian Journal of Program Evaluation. 18(1), 1-23.
- McNamara, G., Joyce, P. & O'Hara, J. (2010). Evaluation of Adult Education and Training Programs. Elsevier, 548-554.
- Meghe, B., Bhise, P. V. & Muley, A. (2013). Evaluation of Training and Development Practices of CTPS using Kirkpatrick Method: A Case Study. International Journal of Application or Innovation in Engineering & Management (IJAIEM), ISSN 2319 – 4847.
- Monaco, E.J. (2014). A Tribute to the Legacy of Donald Kirkpatrick. PDP Communique, 33. Retrieved from <http://www.pdp.albany.edu/>
- Nelson, B. & Dailey, P. (1999). Four Steps for Evaluating Recognition Programs. Workforce, 78(2), 74-78.
- Ornstein, A.C. & Hunkins, F.P. (1998). Curriculum: Foundations, principles and issues. Englawood Cliffs: NJ, Prentice Hall.
- Owen, J.M. (1999). Program evaluation: Forms and Approaches. New York: Routledge.
- Peute, L.W.P. (2013). Human factors methods in health information systems' design and evaluation: The road to success?. Doctorate Thesis. University of Amsterdam.
- Schuman, P.L., Anderson, P.H., Scott, T.W. & Lawton, L. (2001). A Framework for Evaluating Simulations as Educational Tools. Developments in Business Simulation and Experiential Learning, 28, 215-220.
- Stake, R. (1999). Representing Quality in Evaluation. Stake: Quality, 1-7.
- Steele, M.S. (1970). Program Evaluation. A Broader Definition. Journal of Extension, 5-17.
- Topno, H. (2012). Evaluation of Training and Development: An Analysis of Various Models. IOSR Journal of Business and Management (IOSR-JBM), 5(2), 16-22.
- Uşun, S. (2012). Eğitimde Program Değerlendirme. Süreçler Yaklaşımlar ve Modeller. Ankara, Anı Yayıncılık.
- Wall, J. E. (2014). Program Evaluation Model. 9 – Step Process. Sage Solutions.
- W.K. Kellogg Foundation. (2004). Logic Model Development Guide. Michigan.
- Yüksel, İ. & Sağlam, M. (2014). Eğitimde Program Değerlendirme. Yaklaşımlar – Modeller – Standartlar. Ankara, Pegem Akademi Yayıncılık.